Day 7 Cell Ranger Count Worksheet

Author: Chris Ozeroff, July 2024

fastq file —**Cell Ranger**→ filtered count matrix

In this tutorial, you will be taking a single cell RNA-sequencing dataset and map it using the Cell Ranger pipeline. Cell Ranger performs alignment, filtering, and unique molecular identifier and barcode counting. It then outputs several files, including a count matrix which we can analyze using a software package called Seurat.

*Because this can take a while to run, and because we are analyzing such a large dataset, you will get Cell Ranger running on a fastq that is ~10% the size of the actual fastq file from the paper. Then, you will proceed to the next worksheet which will point you to a finished count matrix to analyze in Seurat.

1. Log onto the AWS, git pull from github, and navigate to the srworkshop/projectA/day7 directory. mkdir a directory called cellranger_count inside the day7 directory.

2. You will edit a sbatch script called cellrangerCount_sbatch . rsync this script from the scripts directory into the cellranger_count directory you have just made

3. Open cellrangerCount_sbatch in vim. You will need to edit the error and output file path, the path to the transcriptome directory, and the path to the fastq directory. You will also need to set the ntasks AND local cores equal to 8.

   NOTE: This is the number of cores we have available to us on the AWS. When mapping on your home university supercomputer you can use more.

4. Here is the path to transcriptome directory and the path to the fastq's directory

```
cellranger count --id=T21BM_male19 \
 --fastqs=/scratch/Shares/public/sread2024/data_files/day7a/fastq/sampled_fastq \
 --transcriptome=/scratch/Shares/public/sread2024/cookingShow/day7a/genomes/refdata-gex-GRCh38-2020-A \
```

5. Just a couple more notes about the script. The `--id=T21BM_male19` command will tell cell ranger to put its output files in a directory named "T21BM_male19". The `--sample=DSOX19_1` command gives cell ranger the sample prefix associated with the fastq files for that sample.

6. Now, run the sbatch script

   ```
   -bash-4.2$ sbatch cellrangerCount_sbatch
   ```

7. Check and see if the job has been running for a minute or so (no errors)

```
-bash-4.2$  squeue -u ozeroff

[ozeroff@ip-172-31-29-230 cellranger_count]$ squeue
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
              1148     short cellrang  ozeroff  R       0:36      1 short-dy-m52xlarge-short-5
```

8. At this point you can move onto the next worksheet.

NOTE: Just in case you're interested, when the job is finished (~2hrs), you will have an output directory labeled "T21BM_male19". A link to the overview of all the cellranger outputs is here:

https://www.10xgenomics.com/support/software/cell-ranger/latest/analysis/outputs/cr-outputs-overview

The output file that you'd be most interested in is probably "`filtered_feature_bc_matrix.h5`" in the outs sub-directory. This is a filtered count matrix that we could analyze using a package like Seurat.